

CLAIMS:

1. A method of bandwidth allocation for delivery of stored digital content from at least one server device to at least one client device by way of a network, the method comprising the steps of:

- 5                   a)    prescribing a control variable which represents a target flow rate from the server device to each client device;
- b)    determining time-varying constraints on the flow rate of the content;
- 10                  c)    determining a cost function of the control variables for all clients; and
- d)    prescribing bandwidth to all clients based upon the value of the control variables that maximize the cost function.

2. The method of Claim 1 wherein step (d) comprises performing periodic computations to update the value of the control variable such that the  
15   bandwidth can be continuously allocated to each client.

3. The method of Claim 2 wherein a new client is accepted by:

                  i)    determining an admission capacity of the bandwidth;

                  ii)   admitting a prospective client if the clients minimum allowed value of the control variable is less than the admission capacity; and

20                  iii)   wherein a client admitted for service is guaranteed to have sufficient content flow over the entire session.

4. The method of Claim 3 wherein:

The admission capacity equals a server swing capacity reduced by a predetermined safety margin which may be zero, the swing capacity equaling the difference between the server flow capacity and the sum of the minimum allowed flow rate for all clients.

5. The method of Claim 4 wherein the minimum allowed value of the control variable is initialized to the average consumption rate.

6. The method of Claim 4 wherein the minimum allowed value of the control variable is initialized to the average consumption rate multiplied by a factor greater than one.

7. The method of Claim 4 wherein the server flow capacity varies in a step-wise function over time according to a predetermined schedule and the method of accepting a new client further comprises:

i) determining a sequence of future step changes of server flow capacity;

ii) determining at each time in the sequence of future step changes the value for a future worst swing case capacity, the worst case swing capacity being obtained from the server flow capacity at the time by subtracting an extrapolated present minimum allowed flow rate for all active clients that are potentially active in the future;

iii) admitting a prospective client if a minimum allowed value of the control variable is less than the capacity obtained from the sequence and the present capacity;

iv) whereby a new client can be admitted without compromising the service to existing clients in the presence of previously scheduled server flow capacity changes and all clients accumulate content in their buffers to the extent of their respective buffer capacity.

5           8.     The method according to Claim 7 wherein the method of accepting a new client further comprises:

i)     partitioning the available bandwidth into partitions representing a prescribed service charge;

10           ii)    determining a maximum-minimum capacity for a service charge to be the maximum allowed sum of all minimum allowed flow rates for all clients incurring an equal or lesser service charge, the maximum-minimum capacity equaling the sum of bandwidth allocations for all service charges less than or equal to the service charge;

15           iii)   determinating a service-charge-specific swing capacity for a given service charge as the difference between the maximum-minimum capacity for the service charge and the sum of all minimum allowed flow rates for all clients incurring the service charge or a lesser service charge;

20           iv)    determining the server swing capacity to be the server flow capacity minus a margin and minus the sum of the minimum allowed flow rate for all clients; and

v)     determining the service-charge-specific capacity to be the least of the server swing capacity and the charge specific swing capacity.

9. The method of Claim 8 wherein the method of accepting a new client further comprises:

i) determining the sequence of future times of step changes to server flow capacity over a proposed consumption period;

5 ii) determining for each time in the sequence of future times, the future worst-case value for the service-charge-specific capacity, the worst case value being obtained by using the extrapolated present minimum allowed flow rate of all presently active clients that are potentially active at the time in the future; and

10 iii) admitting a client incurring a given service charge if the minimum flow rate of the client is less than the capacity obtained from the sequence and the present service-charge-specific capacity.

10. The method of Claim 3 wherein step (b) comprises determining time-varying constraints wherein:

15 i) an aggregate flow rate for all clients does not exceed a predetermined server flow capacity;

ii) the flow rate from the server to the client does not exceed a maximum allowed flow rate for the client;

20 iii) the flow rate to the client will never cause a buffer of the client to overflow;

iv) the flow rate to the client stops when the content is exhausted; and

v) the flow rate from the sever will never be less than the client's minimum allowed flow rate which may vary over time and may be zero.

11. The method of Claim 10 wherein the maximum allowed flow rate to the client is given by the minimum of one selected from the following:

- i) a client flow rate ceiling;
- ii) the flow rate required to fill the client buffer before the next periodic computation; and
- iii) the flow rate required to complete the delivery of the content before the next periodic computation.

12. The method of Claim 11 wherein:

- i) the minimum allowed client flow rate cannot increase over time, but may decrease such that the initial value of the control rate is no less than the average consumption rate;
- ii) the control variable for the client is always greater than or equal to the current minimum allowed client flow rate; and
- iii) the client buffer will never underflow until the content is fully consumed at the average consumption rate.

13. The method of Claim 12 wherein the minimum client flow rate is the product of the constant content flow rate from the client to the server that causes the last piece of content to be delivered at the last possible moment as the client buffer is being drained at the average consumption rate, with a factor greater or equal to one.

14. The method of Claim 13 wherein the minimum client flow rate is initialized to the average consumption rate.

15. The method of Claim 13 wherein the minimum client flow rate is initialized to the average consumption rate multiplied by a factor greater than one.

16. The method of Claim 13 further comprising the step of calculating the minimum client flow rate periodically.

5 17. The method of Claim 1 wherein the cost function represents an aggregate flow rate and is the sum of the flow rates for all clients.

18. The method of Claim 1 wherein the cost function represents a service charge and is the sum of the client flow rates multiplied by the client's cost of service.

10 19. The method of Claim 18 wherein:  
step (a) comprises:

i) determining a maximum allowed flow rate and a minimum allowed flow rate for each client;

15 ii) determining a flow rate range for each client as the difference between the maximum allowed flow rate and the minimum allowed flow rate; and

iii) initializing a current flow rate for each client as the minimum allowed flow rate and summing the flow rate into the total server flow rate; and

20 step (d) comprises:

i) computing remaining server bandwidth as the difference between the maximum server flow capacity and the total server flow rate; and

ii) allocating remaining server bandwidth to remaining clients wherein all clients with a given cost of service receive bandwidth equally to the extent of their respective ranges, and no client with a given cost of service being allocated bandwidth unless all clients with a higher cost of service have received their maximum possible allocation of bandwidth.

5

20. The method of Claim 17 wherein step (a) prescribing the control variable comprises:

i) forming a hyperplane of control variables that aggregate to the maximum allowed aggregate flow rate;

10

ii) computing a first vector of minimum allowed client flow rates;

iii) computing a second vector of maximum allowed client flow rates;

15

iv) determining a multi-dimensional shape of which an interior diagonal spans the first and second vectors; and

v) finding a solution of control variables at the point of intersection between the diagonal and the hyperplane.

21. The method of Claim 20 wherein step (v) comprises finding a solution of control variables at the second vector if there is no point of intersection between the diagonal and the hyperplane.

20

22. The method of Claim 17 wherein step (a) comprises:

i) computing the sum of the minimum allowed flow rates to all active clients;

5 ii) computing the sum of the flow rate ranges for all active clients;

iii) computing the difference between the maximum allowed aggregate flow rate and the sum of the minimum;

iv) computing a factor as the ratio of the difference over the sum of the ranges;

10 v) if the factor exceeds 1, the control variable is set to the maximum allowed client flow rate; and

vi) if the factor is less than 1, computing the control variable by multiplying the client range by the factor and then adding the result to the minimum allowed client flow rate.

15 23. The method of Claim 17 wherein:

step (a) comprises:

i) determining the maximum allowed flow rate and minimum allowed flow rate for each client;

20 ii) determining the flow rate range for each client as the difference between the maximum flow rate and the minimum flow rate; and

iii) initializing the control variable for each client as said minimum allowed flow rate and summing the flow rate into a total server flow rate; and

25 step (d) comprises:



i) computing remaining server bandwidth as the difference between the maximum server flow capacity and the total server flow rate; and

ii) allocating remaining server bandwidth to remaining clients to the extent of their respective ranges.

24. The method of Claim 23 wherein the step of allocating remaining server bandwidth further comprises the clients receiving the bandwidth equally to the extent of their respective ranges.

25. The method of Claim 17 wherein the maximized value of the control variable can be determined by:

i) determining a maximum and minimum flow rate for each client;

ii) determining a range between the maximum and minimum flow rates in order to find a flow rate range;

iii) determining available bandwidth by finding the difference between an aggregate flow capacity and a sum of the minimum flow rate;

iv) determining a flow variable by dividing the unused bandwidth by the flow rate range;

v) prescribing the control variable to be the minimum flow rate added to the flow rate range if the flow variable is less than one or prescribing the control variable to be the minimum flow rate added to the flow rate range corrected by the flow variable if the flow favorable is greater than one such that the control variable is easily calculated for each client.

26. The method of Claim 17 wherein at least one session is a session bundle and:

1) the minimum flow constraint is computed as the sum of the minimum flow rate constraints for all bundled sessions;

5 2) the maximum flow constraint is the least of 1) the sum of the maximum individual flow rate constraints of each session and 2) the flow capacity of the channel shared among the bundled sessions; and

3) the target flow rate is computed for the session is apportioned among the bundled constituent sessions.

10 27. The method of Claim 26 wherein the bundled session is an advertisement.

28. A system for allocating bandwidth between a server device and at least one client device, the system comprising:

15 a call acceptance module operative to receive an incoming request for service;

a flow regulator configured to deliver content at a modulated target flow rate, the content being delivered between the server device and a respective client device when a call is accepted by the call acceptance module; and

20 a flow optimizer configured to modulate the target flow rate of the flow regulator in order to optimize the aggregate flow of content.

29. The system of Claim 28 wherein the flow optimizer is configured to modulate the flow rate of the flow regulator in order to optimize charges for the content.

5                    32.    The system of Claim 31 wherein the flow optimizer is configured to generate a cost function of the control variable that corresponds to a maximized value of the control variable.

33. The system of Claim 28 wherein the flow regulator is configured to deliver the modulated flow rate of content in response to the control variable.

- i) determining a maximum and minimum flow rate for each client;

ii) determining a range between the maximum and minimum flow rates in order to find a flow rate range;

iii) determining unused bandwidth by finding the difference between an aggregate flow capacity and a sum of the minimum flow rate;

- iv) determining a flow variable by dividing the unused bandwidth by the flow rate range;

v) prescribing the control variable to be the minimum flow rate added to the flow rate range if the flow variable is less than one or prescribing the control variable to be the minimum flow rate added to the flow rate range corrected by the flow variable if the flow favorable is

greater than one such that the control variable is easily calculated for each client.

35. The system of Claim 33 wherein the call acceptance module is configured to accept a call based upon the value of the control variable.

5           36. A method of bandwidth allocation for delivery of stored digital content from at least one server device to at least one client device by way of a network, the method comprising the steps of:

10               a) prescribing a control variable which represents a target flow rate from the server to each client device based upon the amount of buffer of the client;

              b) determining time-varying constraints on the flow rate of the content;

              c) determining a cost function of the control variables for all clients in response to the size of the client's buffer; and

15               d) prescribing bandwidth to all clients based upon the value of the control variables that maximized the cost function.

37. A method for apportioning a given total flow rate (bandwidth) among a plurality of sessions for clients, the method comprising the steps of:

20               a) computing a sum of minimum flow rate constraints for each session;

              b) computing a sum of flow rate ranges for all active clients;

              c) computing a difference between the given total flow rate and a sum of the flow rate minima;

d) computing a factor as the ratio of the difference over the sum of the flow rate ranges;

e) if the factor exceeds 1, setting the flow rate to the maximum allowed client flow rate; and

5 f) if the factor is less than 1, computing the flow rate by multiplying the client range by the factor and then adding the result to the minimum allowed client flow rate.

10 38. The method of Claim 37 wherein the clients are distinguished by their cost of service and bandwidth is apportioned to clients iteratively, the amount apportioned to each client in a cost category is the amount remaining from prior allocations to clients in higher cost categories.